11638–11647 *Nucleic Acids Research, 2012, Vol. 40, No. 22* *Published online 15 October 2012*
doi:10.1093/nar/gks952

Defining the functional footprint for recognition and repair of deaminated DNA

Michael R. Baldwin and Patrick J. O'Brien*

Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109, USA

Received July 11, 2012; Revised and Accepted September 20, 2012

ABSTRACT

Spontaneous deamination of DNA is mutagenic, if it is not repaired by the base excision repair (BER) pathway. Crystallographic data suggest that each BER enzyme has a compact DNA binding site. However, these structures lack information about poorly ordered termini, and the energetic contributions of specific protein–DNA contacts cannot be inferred. Furthermore, these structures do not reveal how DNA repair intermediates are passed between enzyme active sites. We used a functional footprinting approach to define the binding sites of the first two enzymes of the human BER pathway for the repair of deaminated purines, alkyladenine DNA glycosylase (AAG) and AP endonuclease (APE1). Although the functional footprint for full-length AAG is explained by crystal structures of truncated AAG, the footprint for full-length APE1 indicates a much larger binding site than is observed in crystal structures. AAG turnover is stimulated in the presence of APE1, indicating rapid exchange of AAG and APE1 at the abasic site produced by the AAG reaction. The coordinated reaction does not require an extended footprint, suggesting that each enzyme engages the site independently. Functional footprinting provides unique information relative to traditional footprinting approaches and is generally applicable to any DNA modifying enzyme or system of enzymes.

INTRODUCTION

Tens of thousands of damaged nucleobases are formed each day in a typical human cell and most are repaired by the base excision repair (BER) pathway (1). Spontaneous deamination is one common form of damage that affects pyrimidines and purines (2–4). Although humans have four different DNA glycosylases that recognize deaminated pyrimidines, alkyladenine DNA glycosylase (AAG; also known as methylpurine DNA glycosylase) is the main glycosylase for removal of deaminated purines (hypoxanthine, xanthine and oxanine) (5–8). AAG binds to and flips out the damaged nucleotide and catalyses the hydrolysis of the *N*-glycosidic bond to release the damaged base and create an abasic site. This abasic site is the substrate for AP endonuclease I (APE1) that creates a nick with a 2′-deoxyribose 5′-phosphate group. Subsequently, the 2′-deoxyribose 5′-phosphate is removed by a lyase, the gap is filled by a DNA polymerase and the nick is sealed by a DNA ligase. Work during the past three decades has identified many different substrates of these enzymes and has provided detailed structural information for most individual enzymes. Nevertheless, the energetics of substrate recognition and the means of coordinating these repair pathways remain open questions for BER enzymes.

Crystal structures of AAG in complex with etheno adducts or a pyrrolidine transition state mimic provide a wealth of information regarding close contacts between the protein and the DNA (9–11). These structures guide the development and evaluation of specific hypotheses, but it is not possible to discern the energetic contributions of the binding interactions that are identified. Although AAG has not been crystallized with deoxyinosine-containing DNA, and seems to bind this form of damage relatively weakly, the large catalytic proficiency of 3×10^{17} indicates that this is a preferred substrate (12). The efficient glycosylase reaction and poor ground state binding of deoxyinosine-DNA by AAG was reconciled by the finding that there is an unfavourable equilibrium for nucleotide flipping (12). There are no structures available for AAG bound to B-form undamaged DNA, although recent structures give a glimpse into the binding of AAG to frayed ends of a mismatched duplex (13).

After base excision, AAG binds tightly to the abasic product, and its dissociation is accelerated by APE1 (14). It is intriguing that APE1 seems to bind to essentially the same site as AAG with greater numbers of close protein–DNA contacts occurring downstream of the lesion (15). It has been proposed that there is a transient complex that facilitates the handoff between AAG and APE1, and similar models have been invoked for other

*To whom correspondence should be addressed. Tel: +1 734 647 5821; Fax: +1 734 764 3509; Email: pjobrien@umich.edu

BER glycosylases, including thymine DNA glycosylase (16–19). As it has not been possible to isolate stable complexes of any of these proteins, we have adopted a functional approach to interrogate the properties of this putative complex.

A notable feature of AAG and APE1 is that they contain poorly ordered amino termini that have been refractory to structural studies. This is a common feature of eukaryotic proteins, and such disordered regions are implicated in a wide variety of biological functions (20–22). For AAG, the first 79 amino acids were removed to grow crystals. This truncated protein has full catalytic activity and only slightly decreased DNA binding affinity (9,23). For APE1, the first 35 amino acids were removed for crystallography, and the crystal structure of the full-length protein does not show detectable density for this amino terminus (15,24). Limited proteolysis shows that the amino terminus is accessible (25), and detectable endonuclease activity remains on deletion of the first 60 amino acids (26).

We have developed a functional approach to evaluate the available structural data, and to identify potential contributions from the poorly ordered amino termini of AAG and APE1. We characterized the kinetics of these enzymes using a series of short DNA oligomers in which the location of the damaged site was varied systematically in its separation from a tetranucloetide hairpin. The use of a hairpin circumvents complications caused by fraying of duplex ends. This allowed us to define the limits of the minimal binding sites for efficient catalysis by the full-length proteins, which we have defined as the functional footprints.

MATERIALS AND METHODS

Recombinant proteins

Full-length human AAG and APE1 proteins were expressed in *Escherichia coli* and were purified as previously described (27). The concentration of AAG was determined by active site titration (14), and the concentration of APE1 was determined by absorbance at 280 nm ($E_{280} = 5.6 \times 10^4 M^{-1}cm^{-1}$).

DNA substrates

2′-Deoxyinosine-containing DNA substrates were synthesized by Integrated DNA Technologies or the Keck Facility at Yale and were purified through denaturing polyacrylamide gel electrophoresis (PAGE) (27). Sequences are provided in Supplementary Figure S1. Single-strand DNA concentrations were determined from the absorbance at 260 nm using calculated extinction coefficients. Duplex (hairpin) DNA concentrations were determined from absorbance at 495 nm using the extinction coefficient of fluorescein ($7.5 \times 10^5 M^{-1} cm^{-1}$). These concentrations were confirmed by active site titrations with AAG (27); corrections were <30% in all cases. Duplexes were annealed with 1.5-fold excess complement by heating to 95°C for 5 min and then cooling to 4°C for ~15 min. Hairpins were annealed by heating to 95°C for 10 min and then transferring them to an ice bath. Abasic substrates were prepared by incubating each substrate with AAG under single-turnover conditions, followed by phenol–chloroform extraction and desalting (27). The conversion to abasic sites was verified by alkaline hydrolysis of the abasic sites and analysis by denaturing PAGE. These were stable for months when stored at 4°C and pH 6.5.

General glycosylase assay

Fluorescein-labelled oligonucleotide substrates were incubated with AAG at 37°C in a reaction buffer containing 50 mM of 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid pH 7.0, 1 mM of dithiothreitol, 1 mM of ethylenediaminetetraaceticacid, 0.1 mg/ml of bovine serum albumin, 10% glycerol and sufficient NaCl to reach an ionic strength of 42 mM, unless otherwise specified. Time points were taken by removing a small volume from the reaction and quenching in two volumes of 0.3 M NaOH. Abasic sites were converted to single-strand breaks by heating at 70°C for 15 min. Samples were analysed by denaturing PAGE as described previously (27). Bands were quantified using ImageQuant TL software (GE Healthcare). Reaction time courses were fit using Kaleidagraph software.

Single-turnover glycosylase assay

Reactions were conducted in the reaction buffer, with enzyme present in excess of substrate. Typical reaction conditions included 50 nM of substrate and 200 nM–1 μM AAG. Time courses were continued until reactions reached completion (>7 half-lives). Reaction progress curves were fit by single exponentials (Equation 1). We confirmed that the saturating single-turnover rate constants (k_{max}) were measured by varying the concentration of AAG by at least 4-fold. In all cases, the reaction rate was independent of the concentration of AAG.

$$F = A(1 - \exp(-k_{obs}t)) \quad (1)$$

Multiple-turnover glycosylase assay

AAG multiple-turnover activity in the presence and absence of APE1 was measured with substrate concentration in excess of AAG. Typical substrate concentration was 1 μM, and typical AAG concentration was 25 nM. All reactions contained 0.1 mg/ml of bovine serum albumin that was previously shown to stabilize AAG and avoid artifacts of added proteins on AAG stability (27). When APE1 was present, its concentration was 2 μM. Under these conditions, the initial rates were determined by following reactions to ~20% substrate depletion. Reaction progress curves were linear, indicating that product inhibition was negligible. At saturating substrate concentration, either base excision (k_{max}) or dissociation from the abasic product (k_{off}) can contribute to the overall rate constant (k_{cat}). The relationship of these rate constants is given by Equation 2. This equation can be

rearranged to solve for k_{off} using the measured values of k_{cat} and k_{max} (Equation 3).

$$1/k_{cat} = 1/k_{max} + 1/k_{off} \quad (2)$$

$$k_{off} = k_{max} \times k_{cat}/(k_{max} - k_{cat}) \quad (3)$$

Competition assays to determine relative k_{cat}/K_M values for AAG

Relative specificity for AAG substrates was determined by competition assays. The substrate 35I22 was used as a reference (Supplementary Figure S1). Each hairpin substrate was mixed with the reference substrate at a final DNA concentration of 1 μM; typically, each substrate was present at an equal concentration. The standard reaction buffer was supplemented with NaCl added to achieve a final ionic strength of 150 mM. AAG was added to a final concentration of 25 nM, and the glycosylase activity was followed up to 20% depletion of the reference substrate. Initial velocities are proportional to the relative k_{cat}/K_m values as described in Equation 4 (28).

$$V_A/V_B = ([A] \times (k_{cat}/K_M)_A)/([B] \times (k_{cat}/K_M)_B) \quad (4)$$

The quantum yield of the fluorescein label was not sensitive to the adjacent sequence (data not shown). The ratio of the concentration of hairpin to reference substrate observed directly by gel quantitation, and this led to small corrections (≤20%) in the concentrations of the competing substrates.

Competition assays to determine relative k_{cat}/K_M values for APE1

The general strategy for these experiments was the same as for the AAG competition assays. Abasic DNA was prepared from AAG substrates as described previously. The reaction buffer was supplemented with 3 mM of $MgCl_2$ to give a final free Mg^{2+} concentration of 2 mM, and NaCl was added to reach an ionic strength of 150 mM. Time points were taken by removing 2 μl from a reaction and quenching in 2 μl of 20 mM ethylenediaminetetraaceticacid. Remaining abasic sites were reduced by adding 1 μl of 500 mM $NaBH_4$ in 10 mM NaOH and incubating at room temperature for 30 min. Addition of 1 μl of 85 mM acetic acid was followed by another 30 min incubation, and then 8 μl of formamide loading solution with 5 mM of NaOH was added. Control reactions without borohydride treatment gave identical rates, but in many cases, we observed that abasic DNA was sticking in the wells, and we attributed this to reaction between the abasic aldehyde and the gel. We never observed sticking of DNA that did not contain an abasic site. Typical reactions contained 125 nM each of reference and hairpin substrate and 0.1–2 nM of APE1. In some cases, the hairpin concentration was raised to as much as 500 nM to reach a reaction rate that could be measured in the time that it took APE1 to consume 20% of the reference substrate. Initial rates were measured, and relative specificities were calculated from Equation 4. The relative specificities of poor substrates could only be measured by competing them against intermediately poor substrates (e.g. HP-3 with HP+3). In these cases, the specificity relative to the reference substrate was determined using Equation 5.

$$(k_{cat}/K_M)_A/(k_{cat}/K_M)_C = (k_{cat}/K_M)_A/(k_{cat}/K_M)_B \times (k_{cat}/K_M)_B/(k_{cat}/K_M)_C \quad (5)$$

Energetic contribution of flanking DNA

The contribution of the flanking duplex is given by the difference in free energy between a shortened DNA and an optimal DNA according to Equation 6, in which k_1 and k_2 are the two rate constants being compared, R is the gas constant and T is 310 K.

$$\Delta\Delta G = -RT\ln(k_1/k_2) \quad (6)$$

RESULTS

Outline of the functional footprinting approach

A series of hairpin substrates were prepared to investigate the contribution of upstream and downstream flanking duplex to AAG and APE1 catalysis. Each contained a single deoxyinosine lesion (I•T). The end closest to the lesion was stabilized with a 5′-GAAA hairpin to avoid problems with fraying of DNA ends, and the local sequence context was maintained as much as possible (Figure 1 and Supplementary Figure S1). By comparing the rate constants for glycosylase and AP endonuclease activity at sites located at different distances from the

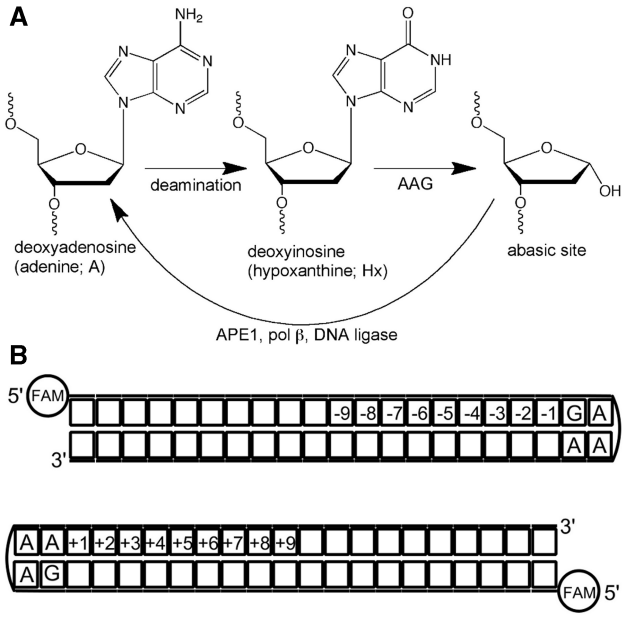

Figure 1. Repair of deoxyinosine by human BER. (**A**) Deaminated purines are excised by AAG to create an abasic DNA repair intermediate that is the substrate for APE1. (**B**) DNA hairpins were 5′-labelled with fluorescein (FAM). Numbers indicate the name for the substrate in which the indicated position contains either deoxyinosine or abasic site that is opposite T.

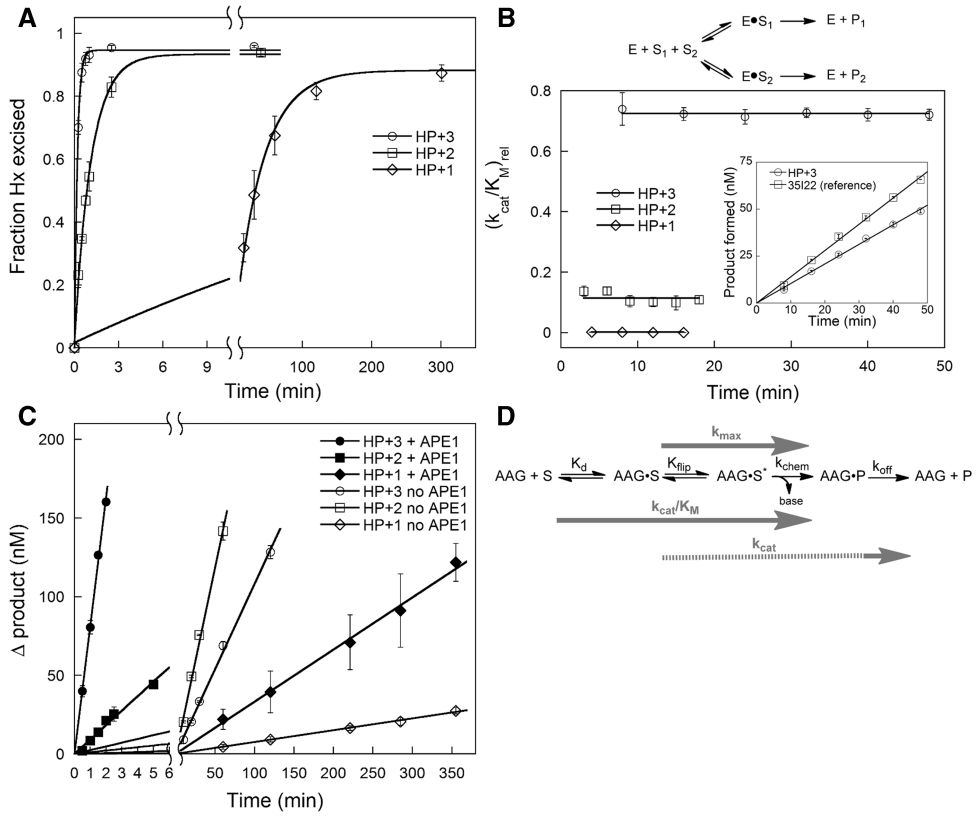

Figure 2. Kinetic characterization of the AAG-catalysed glycosylase reaction. (**A**) Representative single-turnover reactions were fit by single exponentials. (**B**) Representative multiple-turnover competition experiments to determine relative k_{cat}/K_M values. The time course for HP+3 and the reference substrate are shown in the inset, and the ratio of the products yields the relative k_{cat}/K_M value (Equation 4). (**C**) Representative multiple-turnover experiments in the presence (closed symbols) and absence (open symbols) of APE1. The burst phase that occurred for HP+2 and HP+3 before the first time point was subtracted from the data to facilitate comparison of the steady-state rates. The uncorrected data are provided in Supplementary Figure S5. (**D**) The minimal reaction mechanism. AAG must bind to substrate DNA (S) and flip out the damaged base before catalysing *N*-glycosidic bond cleavage. Dissociation from the abasic product (P) is the rate-limiting step at low salt concentration. Grey arrows indicate which steps are monitored by different rate constants. The release of product is considered to be irreversible under the initial rate conditions that were used.

hairpin end, the relative energetic contribution of each base pair can be determined. It was expected that shortening of the upstream and downstream duplex regions would reveal the minimal DNA site required for recognition and repair by each of these BER enzymes. This approach could also be used to probe for the existence of a transient APE1–AAG complex. APE1 is known to stimulate AAG under conditions of excess APE1. If the effect occurs through a protein–protein interaction, the complex may use an extended binding site either upstream or downstream of the lesion, exhibiting a greater functional footprint than either of the individual proteins.

Glycosylase activity of AAG on hairpin oligonucleotides

The minimal kinetic mechanism for the recognition and excision of hypoxanthine (Hx) from an I•T mismatch is depicted in Figure 2D (27). The maximal single-turnover rate constant at saturating amount of enzyme (k_{max}) is limited by the rate of base excision, which includes an unfavourable equilibrium for nucleotide flipping and the *N*-glycosidic bond cleavage step. At low to moderate ionic strength, the rate of reaction with saturating substrate (k_{cat}) is limited by the dissociation of the abasic product. The rate of dissociation increases with increasing NaCl concentration and is no longer rate-limiting at a high concentration of NaCl (27,29).

We first measured the single-turnover rate constants at saturating concentration of AAG for each of the DNA substrates. In all cases, the single-turnover reaction was saturated and followed a single exponential (representative data are shown in Figure 2A). The values of k_{max} for each of the substrates are plotted versus the lesion position in Figure 3A (Supplementary Table S1). Remarkably, lesions located only 2 bp away from the hairpin were relatively good substrates by this measure, and lesions that were directly adjacent to the hairpin were only ~100-fold worse than the optimal substrates.

The K_M values for the substrates are too low to be measured directly with our fluorescence-based assay. However, k_{cat}/K_M values can be determined by direct competition even at concentrations above the K_M. The apparent second order rate constant k_{cat}/K_M, commonly referred to as the specificity constant, reports on all of the steps up to and including the first irreversible step, and it

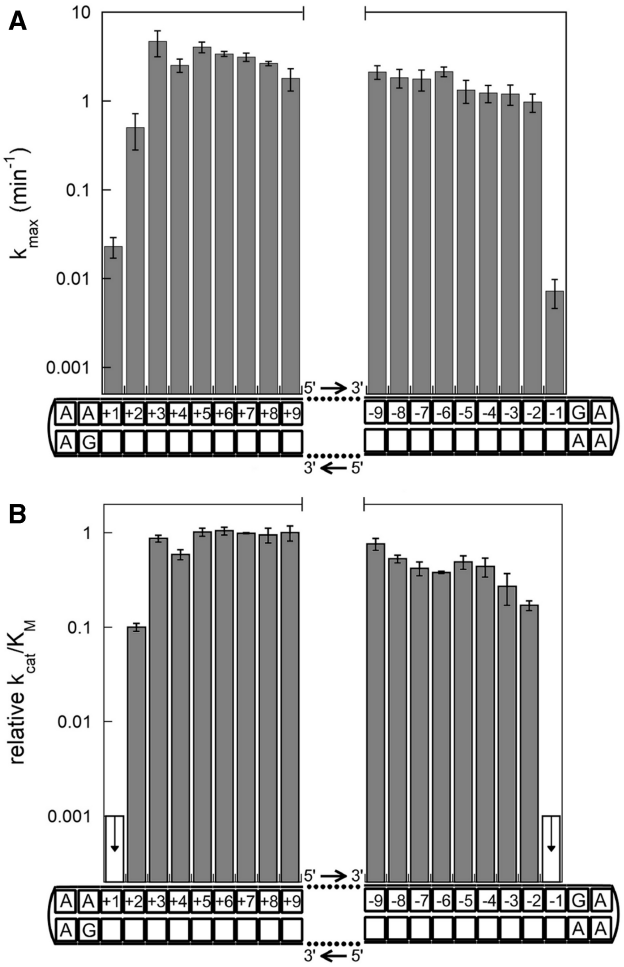

Figure 3. Glycosylase activity of AAG towards hairpin DNA. (**A**) Maximal single-turnover rate constants (k_{max}) for excision of Hx are plotted as a function of lesion position. Differences could be because of changes in nucleotide flipping or *N*-glycosidic bond cleavage. (**B**) The relative specificity constants (k_{cat}/K_M) were determined by competition and are sensitive to any of the steps up to and including *N*-glycosidic bond cleavage. All rate constants were measured in at least three independent experiments, and the error bars indicate the mean ± [standard deviation (SD)]. The open bars for +1 and −1 substrates indicate the upper limit for the k_{cat}/K_M value because no competition was observed.

includes substrate binding and *N*-glycosidic bond hydrolysis. It is important to use a relatively high concentration of salt for these measurements because DNA binding is essentially irreversible at low salt concentration, and there is a high commitment to catalysis (14). At the ionic strength conditions that we used (150 mM), the relative values of k_{cat}/K_M reflect the intrinsic specificity of AAG (Supplementary Figure S6). Representative data are shown in Figure 2B, and the relative k_{cat}/K_M values are plotted in Figure 3B (Supplementary Table S2). Shortening of the upstream region has no significant effect on the k_{cat}/K_M value until the lesion is located 2 bp away from the hairpin, which is decreased by 10-fold, and the position adjacent to the hairpin has a k_{cat}/K_M value that is decreased by >1000-fold relative to the optimal values. Shortening of the downstream DNA caused modest decreases in k_{cat}/K_M, so that the position 2 bp away from the hairpin was only decreased by 5-fold. However, the position immediately adjacent to the hairpin was decreased by >1000-fold relative to the optimal values. The 1000-fold (4 kcal/mol) reduction observed for replacement of either flanking duplex region with a GAAA hairpin is a lower limit for the binding energy that is obtained by AAG because AAG may obtain some favourable binding energy from the hairpin structure (see later in the text). These experiments reveal that AAG has a remarkably compact binding site and that it can productively engage lesions located in close proximity to a DNA hairpin.

Endonuclease activity of APE1 on hairpin oligonucleotides

To obtain the energetic contributions of upstream and downstream flanking regions on the APE1-catalysed hydrolysis of abasic site DNA, we used AAG to convert each of the hairpin oligonucleotides into abasic DNA intermediates. The oligonucleotides were purified and then incubated with APE1 to monitor the AP endonuclease reaction. We did not pursue single-turnover experiments because the APE1-catalysed reaction is too fast to be measured even by rapid quench (30). Nevertheless, we could readily measure the relative k_{cat}/K_M values for this series of substrates through multiple-turnover kinetics by direct competition as described for AAG (Supplementary Figure S4). These experiments were carried out with 150 mM of NaCl because we found that the relative k_{cat}/K_M values were independent of ionic strength of >100 mM (Supplementary Figure S7). At 50 mM ionic strength, DNA binding to APE1 seems to be irreversible because there is no discrimination between substrates (14). The resulting data reveal that APE1 has a substantially larger footprint than AAG (Figure 4 and Supplementary Table S2).

As the abasic site is moved closer to the upstream hairpin, a detectable drop in activity is first observed at the position 4 bp away from the hairpin (+4). For each additional base pair that the abasic site is shifted towards the hairpin, an additional 10–100-fold drop in activity was observed. No activity was detected when the abasic site was positioned immediately adjacent to the hairpin (+1), placing a lower limit of 10^6-fold contribution (8.5 kcal/mol) of the upstream DNA binding site to APE1 binding and catalysis. Less dramatic effects were observed for shortening the DNA downstream of the abasic site, but nevertheless, a significant (3-fold) reduction in k_{cat}/K_M was observed between positions −6 and −5. The value of k_{cat}/K_M continued to drop for each base pair that was removed from the downstream DNA, and the abasic site that was immediately adjacent to the downstream hairpin (−1) is ~1000-fold less efficient of a substrate than an optimal internal abasic site. Thus, the downstream DNA binding interactions contribute at least 4 kcal/mol to binding and catalysis by APE1.

Stimulation of AAG dissociation by APE1

We previously found that APE1 efficiently displaces AAG and stimulates its multiple-turnover glycosylase activity

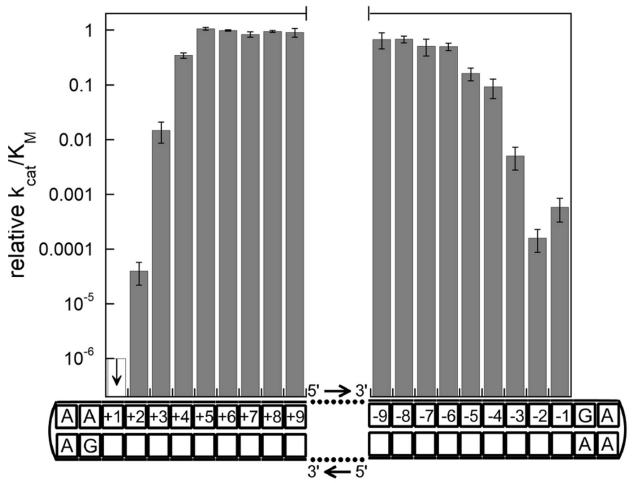

Figure 4. AP endonuclease activity of APE1 towards hairpin DNA. Relative specificity constants (k_{cat}/K_M) were determined by competition and were plotted as a function of lesion position. Rate constants are the average of at least three independent experiments, and the error bars indicate the mean (SD).

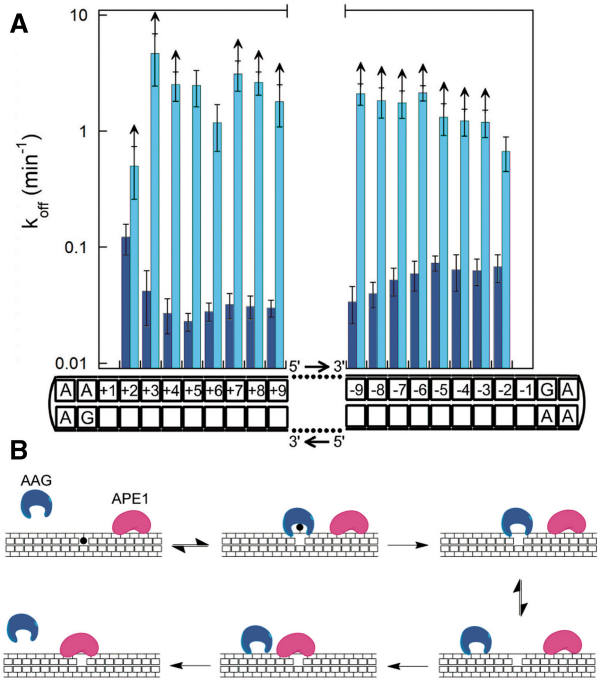

Figure 5. Stimulation of AAG by APE1 on asymmetric hairpins. (**A**) Multiple-turnover glycosylase activity was measured in the presence (light bars) and absence of APE1 (dark bars). The mean (SD) is shown for at least three independent experiments for each substrate. Arrows indicate those substrates for which the stimulated k_{cat} value reaches the rate constant for excision of Hx. (**B**) Model for the displacement of AAG by APE1, whereby AAG leaves the abasic site to allow APE1 to bind. Tight binding by APE1 prevents rebinding of AAG and leaves it free to dissociate more quickly from undamaged DNA.

on substrates with upstream and downstream duplex regions that were as short as 6 bp by increasing the rate of AAG dissociation (14). This observation ruled out the possibility of a functional APE1–AAG complex that uses the full DNA binding surfaces of both proteins, but cannot rule out the possibility that the DNA binding sites are only partially used or partially overlapping. In the current work, we tested the stimulation of AAG by APE1 for substrates that have much shorter upstream and downstream duplex regions. Under the conditions used, the steady-state rate of AAG is limited by the rate of dissociation from the abasic DNA product. Representative data for the stimulation by APE1 are provided in Figure 2C. The dissociation rate constants in the presence and absence of APE1 are summarized in Figure 5 and Supplementary Table S1. The substrates with lesions immediately adjacent to the hairpins were omitted from this plot because the AAG reaction was no longer limited by product release. In many cases, the presence of APE1 accelerated the rate of product release to the point where AAG was limited by the rate of *N*-glycosidic bond cleavage (i.e. $k_{cat} = k_{max}$). In these cases, the observed rate constant is a lower limit for the rate of AAG dissociation (indicated by arrows in Figure 5).

The notable decrease in the APE1-stimulation at the +2 and −2 positions is best explained by decreased affinity of APE1 for binding to these terminal sites (Figure 4). If one considers the substrates that are located within 3 bp from the hairpin that provide optimal binding of AAG (Figure 3B), these all show robust stimulation by APE1. The simplest model to explain the similar footprint for APE1 catalytic activity and for the APE1-stimulation of AAG is that APE1 directly competes with AAG for binding to the abasic site. A transient excursion of AAG can allow access of APE1 to the abasic site, without the requirement for a protein–protein interaction that would decrease the affinity of AAG for DNA.

DISCUSSION

We performed kinetic analyses for a series of homologous substrates in which the lesion position is altered relative to a hairpin to determine the energetic contributions of specific base pairs to productive binding by AAG and APE1. This defines the minimal site required for catalytic recognition. These results are generally consistent with the known crystal structures, but reveal surprising features for DNA binding by AAG and APE1. We also investigated the stimulation of AAG by APE1 and provide evidence that APE1 does not form a complex with AAG on DNA. Instead, it seems that AAG transiently exposes the abasic repair intermediate to allow binding of APE1.

Functional footprint of AAG

Crystal structures of AAG are of truncated protein that lacks the amino terminus (residues 1–79). The amino terminus of AAG is poorly conserved even among mammals (Supplementary Figure S2), and humans have several splice variants that differ in this region. This truncated enzyme retains full glycosylase activity in single-turnover assays (23), but it exhibits an increased rate of dissociation from the abasic DNA product that results in a decreased ability to diffuse along DNA in search for sites of damage (29). The crystal structure of

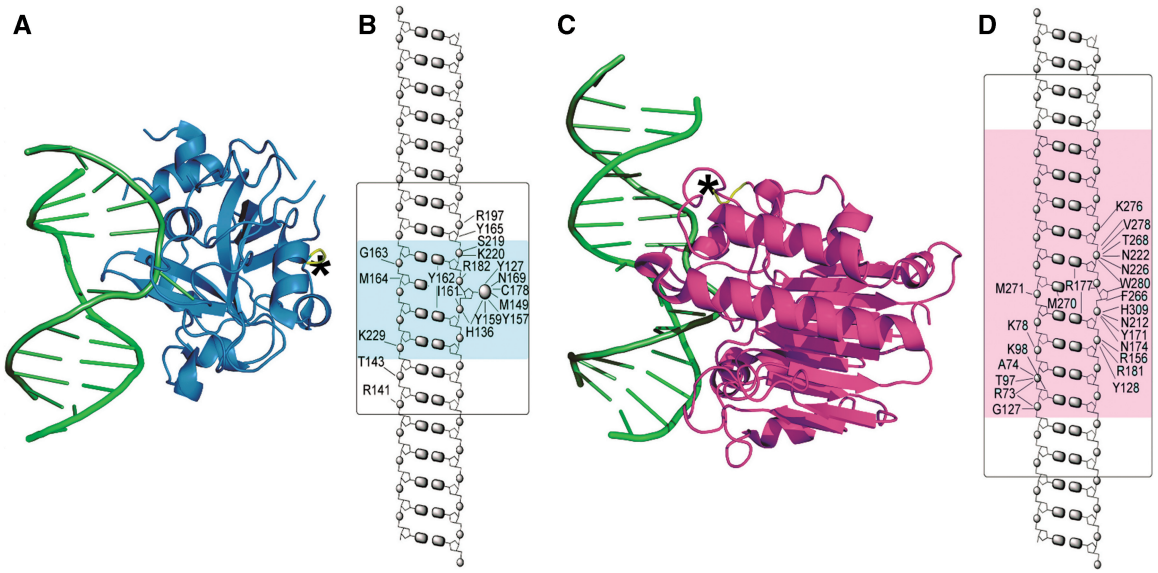

Figure 6. Comparison of crystallographic and functional footprints for AAG and APE1. (**A**) Crystal structure of truncated AAG (residues 80–298) bound to DNA is from the pdb (1EWN) and is displayed using Pymol (10). (**B**) Crystallographic contacts are indicated by an arrow and the amino acid residue number (10). The box indicates the site size required for maximal AAG activity. The boundaries of the box indicate the position of the first base pair that contributes >0.5 kcal/mol to k_{cat}/K_M assuming that AAG recognizes a hairpin as either 0 (shaded) or 2 bp (open) of DNA. (**C**) Crystal structure of truncated APE1 (residues 40–318) bound to abasic DNA is from the pdb (1DEW) and is displayed using Pymol (15). (**D**) Crystallographic contacts for APE1 are from the structure of APE1 bound to an abasic-DNA substrate in the absence of divalent metal ions, and the box is drawn as for AAG (15). The top of the figure is defined as downstream relative to the lesion, and the amino termini is indicated with an asterisk.

truncated AAG bound to 1,N^6-ethenoadenosine-containing DNA shows that the DNA interactions are localized to 8 bp around the damaged nucleotide (10). We compare these crystallographic contacts with the functional footprint for full-length AAG that is defined by the base pairs that contribute to catalytic recognition by AAG (Figure 6B). As it is not possible to predict to what extent the 5′-GAAA hairpin will mimic B-form DNA, we show the two extremes in which AAG either does not interact with the hairpin (dark shade) or does interact with the hairpin (light shade). The strict conservation of R141 that is upstream and R197 that is downstream of the lesion site among AAG homologues suggests that these electrostatic interactions are favourable. As the more conservative footprint (shaded box; Figure 6B) would not involve a contact between the DNA and these conserved arginine residues, it is likely that AAG is able to interact favourably with the phosphate backbone of the hairpin (open box).

Previously, DNase I protection was used to probe the DNA binding footprint of AAG on εA- and inosine-containing DNA (31,32). The resulting footprint is much larger, with up to 11 bp upstream and 5 bp downstream showing some protection from DNase I. This larger protection is observed because the steric bulk of DNase I and AAG preclude the active site of DNase I from closer approach (31,32).

The excellent agreement that is observed between the crystal contacts in the structure of the truncated protein and in the functional interactions with the full-length protein suggests that the amino terminus of AAG does not make direct interactions with the DNA either upstream or downstream of the crystallographic contacts of the catalytic domain. Therefore, the increased affinity of the full-length protein could be because of either direct interactions with the DNA within the identified binding site of 8 bp or conformational changes in the catalytic domain that serve to strengthen the interaction with the DNA.

Functional footprint of APE1

The first crystal structures of APE1 in complex with DNA used a truncated protein that lacked the amino terminus [residues 1–35; (15)]. A crystal structure of the full-length APE1 has also been reported, but no density could be observed for this amino terminal region (24). Nevertheless, the amino terminus is extremely well conserved among mammals (Supplementary Figure S3), and the location of the amino terminus in the crystal structure is in close proximity to the DNA (Figure 6C). The DNA contacts of APE1 that are observed in the crystal structure are localized to the 9 bp region surrounding the damaged site (Figure 6D). As described for AAG, we considered the possibility that APE1 may or may not be able to interact with the 5′-GAAA hairpin. The downstream flanking duplex is important for catalytic recognition of APE1, with significant catalytic contributions derived from binding to the DNA 3–5 bp downstream (Figure 6D). Closer examination of the protein surface in this region reveals several highly conserved positively charged amino acids that could contribute to binding (K224 is 5.3 Å, K227 is 6.9 Å and K228 is 5.9 Å away from phosphate oxygen atoms). It is possible that the catalytic contribution in this region is electrostatic, without

direct contact. Alternatively, the DNA–protein interaction surface may be much more extensive than was observed in the crystal structure. It is worth noting that the DNA makes a crystal contact on this end that provides a possible explanation for why protein–DNA contacts were not observed in this region. Regardless of the origin of this discrepancy between crystal structures and the functional footprint that we have determined, the functional data provide a more complete definition of the APE1 DNA binding site.

The minimal site size of APE1 was previously probed by hydroxyl radical footprinting (33). This study found a similar upstream footprint of 5 bp, but only 3 bp downstream of the abasic site. As hydroxyl radical footprinting requires that the protein–DNA complex show either increased or decreased reactivity, it is expected that this approach often underestimates the full extent of protein–DNA interactions. APE1 interaction with single-strand and duplex DNA was previously characterized by comparing the inhibition of different length oligonucleotides (34,35). Experiments with blunt-ended oligonucleotides concluded that APE1 can bind favourably to 10 bp of duplex DNA (35), which is similar to our results with hairpin DNA. More recently, a protein modification strategy was used to probe the accessible surface area of APE1 when it is bound to DNA (36). Although this approach has the same caveats as DNA modification approaches, it is interesting that these experiments showed protection of K224, K227 and K228 when abasic DNA was bound. These are the residues that are closest to the DNA downstream of the abasic site in the crystal structure. Thus, the kinetic data and the protein accessibility experiments are consistent with the existence of a greater interaction surface than observed in the existing crystal structures.

Mechanism of stimulation of AAG by APE1

The functional footprinting approach of correlating rate constants to shortened DNA substrates can be applied to any kinetic assay; therefore, it is not limited to questions of specificity for a single enzyme. We were interested in the question of how APE1 is able to stimulate the dissociation of AAG. Previously, it was shown that the multiple-turnover glycosylase activity of AAG is limited by the rate of dissociation from the abasic product, and this rate is greatly accelerated in the presence of APE1 (14,27). Displacement of AAG by APE1 does not require APE1 endonuclease activity because robust stimulation is observed in the absence of Mg^{2+}, an essential cofactor for APE1 catalysis. If AAG and APE1 were to form a transient complex, then it is expected that the protein–DNA interface of the two proteins would be greater than the interface of either protein alone.

No evidence for such an interaction was observed (Figure 5). Instead, we observe efficient stimulation of AAG even when the site of damage is immediately adjacent to the hairpin that is placed either upstream or downstream. This indicates that APE1 does not have a preferred orientation for displacing AAG. We suggest that AAG must transiently leave the abasic site and that this makes it accessible to APE1. Once APE1 binds, then the abasic site is no longer available to AAG, and it dissociates more quickly from an undamaged DNA site. In the absence of APE1, it is most likely that AAG will return to the site and remain bound to the abasic product. This dynamic exchange model (Figure 5B) is particularly attractive for explaining how the activities of many different DNA glycosylases are coordinated by APE1 because the glycosylases belong to four different structural families and no common interaction motif has been detected.

General considerations for determining the functional footprint of DNA-binding enzymes

Use of hairpin oligonucleotides is essential to overcoming the deleterious effects of end-fraying. However, the resolution is limited by the interaction between the protein and the hairpin, which could be different from the interaction with normal duplex DNA. For the current study, we used 5′-GAAA tetranucleotide hairpins because these are known to form exceptionally stable structures (37–39). In many cases, it could be advantageous to use non-nucleotide connectors to stabilize the end without introducing additional charge. For example, polyethylene glycol connects are available for commercial DNA synthesis, and these have been shown to effectively stabilize oligonucleotide ends (40,41). We did not use such hairpin structures in the current study because APE1 has robust endonuclease activity towards this type of linkage (14). Comparison of the crystallographic footprint and functional footprint for AAG suggests that this enzyme is able to make favourable interactions with the phosphates of the GAAA nucleotides in the hairpin. The APE1 functional footprint is significantly larger than the crystallographic footprint; therefore, it is not possible to evaluate whether APE1 interacts directly with the hairpin. However, it is expected that electrostatic interactions would make favourable energetic contributions even if the geometry of the hairpin were suboptimal relative to B-form duplex.

We performed a full kinetic characterization for AAG with these hairpin substrates, but this is not necessary to obtain the functional footprint. By using direct competition, the relative k_{cat}/K_M values can be rapidly determined. As k_{cat}/K_M monitors the reaction from free substrate in solution to the transition state for bond cleavage in the enzyme–substrate complex, it provides information about all of the enzyme–substrate interactions. When measuring k_{cat}/K_M, it is critical to ensure that binding is not irreversible, otherwise k_{cat}/K_M may simply reflect the rate of substrate association. Our approach of using high NaCl to increase the dissociation rate constant may be generally applicable to other DNA modifying enzymes because enzymes that interact with DNA commonly use electrostatic interactions with the backbone phosphoryl groups.

The functional footprinting approach has advantages over traditional footprinting or structural approaches. It is quantitative, so that energetic contributions of protein–DNA contacts can be determined (Supplementary

Table S2). There are no restrictions on the protein or proteins that are used. Thus, full-length proteins can be studied, and steady-state kinetics requires small amounts of material. This approach can also be extended to interrogate putative protein–protein complexes on DNA. For many DNA repair pathways, it is postulated that protein–protein interactions facilitate the transition of reaction intermediates from one enzyme to the next. Transient interactions that could not be studied directly by structural methods would be revealed by the requirement for a larger segment of DNA than is required by either protein alone.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figures 1–7 and Supplementary Reference [42].

ACKNOWLEDGEMENTS

The authors thank members of the O'Brien lab for helpful discussions and critical comments on the manuscript.

FUNDING

U.S. National Institutes of Health [CA122254]. Funding for open access charge: NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
2. Myrnes,B., Guddal,P.H. and Krokan,H. (1982) Metabolism of dITP in HeLa cell extracts, incorporation into DNA by isolated nuclei and release of hypoxanthine from DNA by a hypoxanthine-DNA glycosylase activity. *Nucleic Acids Res.*, **10**, 3693–3701.
3. Dianov,G. and Lindahl,T. (1991) Preferential recognition of I.T base-pairs in the initiation of excision-repair by hypoxanthine-DNA glycosylase. *Nucleic Acids Res.*, **19**, 3829–3833.
4. Lindahl,T. and Nyberg,B. (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry*, **13**, 3405–3410.
5. Saparbaev,M. and Laval,J. (1994) Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases. *Proc. Natl Acad. Sci. USA*, **91**, 5873–5877.
6. Hitchcock,T.M., Dong,L., Connor,E.E., Meira,L.B., Samson,L.D., Wyatt,M.D. and Cao,W. (2004) Oxanine DNA glycosylase activity from mammalian alkyladenine glycosylase. *J. Biol. Chem.*, **279**, 38177–38183.
7. Wuenschell,G.E., O'Connor,T.R. and Termini,J. (2003) Stability, miscoding potential, and repair of 2′-deoxyxanthosine in DNA: implications for nitric oxide-induced mutagenesis. *Biochemistry*, **42**, 3608–3616.
8. Dong,L., Meira,L.B., Hazra,T.K., Samson,L.D. and Cao,W. (2008) Oxanine DNA glycosylase activities in mammalian systems. *DNA Repair*, **7**, 128–134.
9. Lau,A.Y., Scharer,O.D., Samson,L., Verdine,G.L. and Ellenberger,T. (1998) Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell*, **95**, 249–258.
10. Lau,A.Y., Wyatt,M.D., Glassner,B.J., Samson,L.D. and Ellenberger,T. (2000) Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl Acad. Sci. USA*, **97**, 13573–13578.
11. Lingaraju,G.M., Davis,C.A., Setser,J.W., Samson,L.D. and Drennan,C.L. (2011) Structural basis for the inhibition of human alkyladenine DNA glycosylase (AAG) by 3,N4-ethenocytosine-containing DNA. *J. Biol. Chem.*, **286**, 13205–13213.
12. O'Brien,P.J. and Ellenberger,T. (2004) Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J. Biol. Chem.*, **279**, 9750–9757.
13. Setser,J.W., Lingaraju,G.M., Davis,C.A., Samson,L.D. and Drennan,C.L. (2012) Searching for DNA lesions: structural evidence for lower- and higher-affinity DNA binding conformations of human alkyladenine DNA glycosylase. *Biochemistry*, **51**, 382–390.
14. Baldwin,M.R. and O'Brien,P.J. (2010) Nonspecific DNA binding and coordination of the first two steps of base excision repair. *Biochemistry*, **49**, 7879–7891.
15. Mol,C.D., Izumi,T., Mitra,S. and Tainer,J.A. (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature*, **403**, 451–456.
16. Fitzgerald,M.E. and Drohat,A.C. (2008) Coordinating the initial steps of base excision repair. Apurinic/apyrimidinic endonuclease 1 actively stimulates thymine DNA glycosylase by disrupting the product complex. *J. Biol. Chem.*, **283**, 32680–32690.
17. Waters,T.R., Gallinari,P., Jiricny,J. and Swann,P.F. (1999) Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J. Biol. Chem.*, **274**, 67–74.
18. Hill,J.W., Hazra,T.K., Izumi,T. and Mitra,S. (2001) Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair. *Nucleic Acids Res.*, **29**, 430–438.
19. Marenstein,D.R., Chan,M.K., Altamirano,A., Basu,A.K., Boorstein,R.J., Cunningham,R.P. and Teebor,G.W. (2003) Substrate specificity of human endonuclease III (hNTH1). *Effect of human APE1 on hNTH1 activity*. *J. Biol. Chem.*, **278**, 9005–9012.
20. Hegde,M.L., Hazra,T.K. and Mitra,S. (2010) Functions of disordered regions in mammalian early base excision repair proteins. *Cell. Mol. Life Sci.*, **67**, 3573–3587.
21. Ward,J.J., Sodhi,J.S., McGuffin,L.J., Buxton,B.F. and Jones,D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.*, **337**, 635–645.
22. Dyson,H.J. and Wright,P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **6**, 197–208.
23. O'Brien,P.J. and Ellenberger,T. (2003) Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines. *Biochemistry*, **42**, 12418–12429.
24. Beernink,P.T., Segelke,B.W., Hadi,M.Z., Erzberger,J.P., Wilson,D.M. 3rd and Rupp,B. (2001) Two divalent metal ions in the active site of a new crystal form of human apurinic/apyrimidinic endonuclease, Ape1: implications for the catalytic mechanism. *J. Mol. Biol.*, **307**, 1023–1034.
25. Strauss,P.R. and Holt,C.M. (1998) Domain mapping of human apurinic/apyrimidinic endonuclease. Structural and functional evidence for a disordered amino terminus and a tight globular carboxyl domain. *J. Biol. Chem.*, **273**, 14435–14441.
26. Izumi,T. and Mitra,S. (1998) Deletion analysis of human AP-endonuclease: minimum sequence required for the endonuclease activity. *Carcinogenesis*, **19**, 525–527.
27. Baldwin,M.R. and O'Brien,P.J. (2009) Human AP endonuclease 1 stimulates multiple-turnover base excision by alkyladenine DNA glycosylase. *Biochemistry*, **48**, 6022–6033.
28. Fersht,A. (1999) *Structure and Mechanism in Protein Science*. W.H. Freeman, New York.
29. Hedglin,M. and O'Brien,P.J. (2008) Human alkyladenine DNA glycosylase employs a processive search for DNA damage. *Biochemistry*, **47**, 11434–11445.

30. Maher,R.L. and Bloom,L.B. (2007) Pre-steady-state kinetic characterization of the AP endonuclease activity of human AP endonuclease 1. *J. Biol. Chem.*, **282**, 30577–30585.
31. Roy,R., Biswas,T., Hazra,T.K., Roy,G., Grabowski,D.T., Izumi,T., Srinivasan,G. and Mitra,S. (1998) Specific interaction of wild-type and truncated mouse N-methylpurine-DNA glycosylase with ethenoadenine-containing DNA. *Biochemistry*, **37**, 580–589.
32. Miao,F., Bouziane,M. and O'Connor,T.R. (1998) Interaction of the recombinant human methylpurine-DNA glycosylase (MPG protein) with oligodeoxyribonucleotides containing either hypoxanthine or abasic sites. *Nucleic Acids Res.*, **26**, 4034–4041.
33. Nguyen,L.H., Barsky,D., Erzberger,J.P. and Wilson,D.M. 3rd. (2000) Mapping the protein-DNA interface and the metal-binding site of the major human apurinic/apyrimidinic endonuclease. *J. Mol. Biol.*, **298**, 447–459.
34. Nevinsky,G.A. (2011) Structural, thermodynamic, and kinetic basis for the activities of some nucleic acid repair enzymes. *J. Mol. Recognit.*, **24**, 656–677.
35. Beloglazova,N.G., Kirpota,O.O., Starostin,K.V., Ishchenko,A.A., Yamkovoy,V.I., Zharkov,D.O., Douglas,K.T. and Nevinsky,G.A. (2004) Thermodynamic, kinetic and structural basis for recognition and repair of abasic sites in DNA by apurinic/apyrimidinic endonuclease from human placenta. *Nucleic Acids Res.*, **32**, 4134–5146.
36. Yu,E., Gaucher,S.P. and Hadi,M.Z. (2010) Probing conformational changes in Ape1 during the progression of base excision repair. *Biochemistry*, **49**, 3786–3796.
37. Hirao,I., Nishimura,Y., Tagawa,Y., Watanabe,K. and Miura,K. (1992) Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAA AGC) and their effect on DNA sequencing. *Nucleic Acids Res.*, **20**, 3891–3896.
38. Santini,G.P., Cognet,J.A., Xu,D., Singarapu,K.K. and Herve du Penhoat,C. (2009) Nucleic acid folding determined by mesoscale modeling and NMR spectroscopy: solution structure of d(GCGA AAGC). *J. Phys. Chem. B*, **113**, 6881–6893.
39. Antao,V.P., Lai,S.Y. and Tinoco,I. Jr. (1991) A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res.*, **19**, 5901–5905.
40. Ng,P.S. and Bergstrom,D.E. (2004) Protein-DNA footprinting by endcapped duplex oligodeoxyribonucleotides. *Nucleic Acids Res.*, **32**, e107.
41. Ng,P.S., Pingle,M.R., Balasundarum,G., Friedman,A., Zu,X. and Bergstrom,D.E. (2003) Endcaps for stabilizing short DNA duplexes. *Nucleosides Nucleotides Nucleic Acids*, **22**, 1635–1637.
42. Goujon,M., McWilliam,H., Li,W., Valentin,F., Squizzato,S., Paern,J. and Lopez,R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.*, **38**, W695–W699.